The Arizona Model for Measuring Educator Effectiveness

Preliminary Findings from Year 1 Pilot Implementation of the Arizona Department of Education Model Educator Evaluation Process

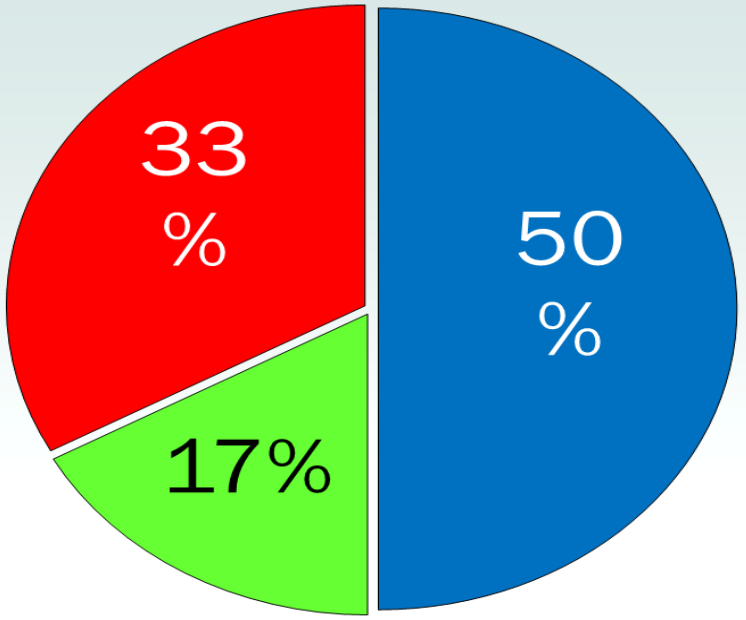
Dr. Karen Butterfield, Associate Superintendent
Todd Petersen, Deputy Associate Superintendent
ADE's Highly Effective Teachers & Leaders Division

Arizona Department of Education

Project Partners

- **Arizona Department of Education (ADE)**
  - Highly Effective Teachers and Leaders Division
  - Research & Evaluation Division
- **West Comprehensive Center (WCC)**
- **Regional Education Laboratory West (REL West)**
- **Five Pilot LEAs:**
  - Williams Unified School District
  - Bisbee Unified School District
  - Maricopa Unified School District
  - Stanfield Elementary School District
  - Accelerated Elementary and Secondary Charter School

Teacher Evaluation Instrument: General Weighting of Three Components



60 Points (50%) = Teaching Performance

40 Points (33%) = Student Academic Progress

20 Points (17%) = Survey Data, Peer Review

120 Points Possible

Teacher Evaluation Instrument: Component Overview

- **Teaching Performance Component – 50% (60 points)**
  - Observation Rubric: Danielson Framework for Teaching (2011)
  - Teachscape platform used to collect observation data
- **Student Academic Progress Component – 33% (40 points)**
  - Rating Tables developed for teachers aligned to their area of instruction and available data
- **Survey Component – 17% (20 points)**
  - Student Survey
  - Parent Survey
  - Peer-Review
  - Self-Review

Year 1 Qualitative Analysis Overview

- Two rounds of focus groups/interviews (February & May)
  - Round 1: 46 teachers, 13 principals
  - Round 2: 35 teachers (5 new), 11 principals (0 new)
- Online surveys emailed to participating teachers in May
  - 165 participating pilot teachers responded
- Mix of grade spans (elementary/secondary) & experience levels represented in focus group/interview, survey data

Data were compiled, coded, and analyzed to identify key findings and themes

Qualitative Analysis: Preliminary Findings & Themes

- Overall Evaluation Model
- Training & Communication
- Time Issues
- Teaching Performance (50%; 60 points)

  Observations/Conferencing
- Surveys: Student, Parent, Peer Review (17%; 20 points)
- Student Academic Progress (33%; 40 points)
- Summative Conferences & Ratings

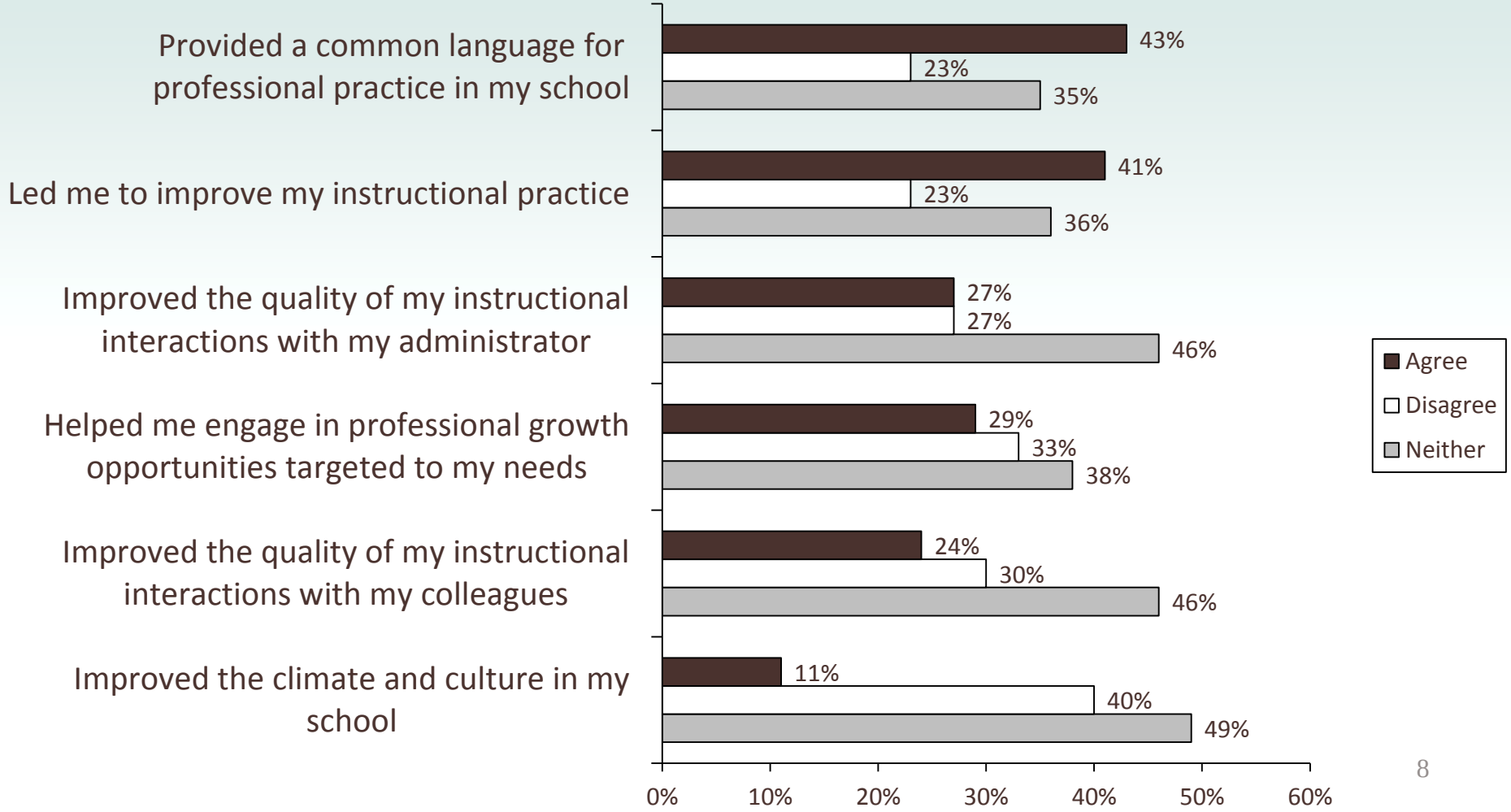
Preliminary Findings:

Overall Evaluation Model

- In interviews, principals called the new system less biased than previous systems & generally believed that it resulted in higher quality feedback for teachers.
- Despite many focus group teachers reporting frustrations with ratings lower than expected:
  - 52% of surveyed teachers agreed that the new teacher evaluation process is fair.
  - 45% of surveyed teachers felt new process is an improvement over their prior teacher evaluations.

Preliminary Results: End-of-Year Surveys of Teachers

The new ADE model teacher evaluation process has…



Preliminary Findings: Training & Communication

- 63% of surveyed teachers agreed that "the criteria on which I was evaluated were made clear to me".
- 55% agreed that their training was adequate for them to effectively participate.
  - Some focus group teachers reported difficulty using *Teachscape* online & hoped for more training on the use of the website.

Preliminary Findings: Training/Communication

- Principals felt well trained to assess Danielson Framework Domains 2 (Classroom Environment) & 3 (Instruction), less prepared to assess Domains 1 (Planning & Preparation) & 4 (Professional Responsibilities).
- Some communication disconnects within LEAs & schools
- Key remaining question areas for participants:
  - Uncertainty around component scoring
  - Equity issues across student populations
  - Among principals: Coaching conversations with teachers

Preliminary Findings:

Time Issues

- All participants cited labor intensiveness, time burdens involved with the new, more thorough observations (e.g., new forms, multiple domains, more evidence).
- Time estimates from principals in one focus group:
  - Approximately 3 hours per observation cycle
  - Round of observations for 20 teachers takes about one month
- *Teachscape* technology helped efficiency, but process is still lengthy.

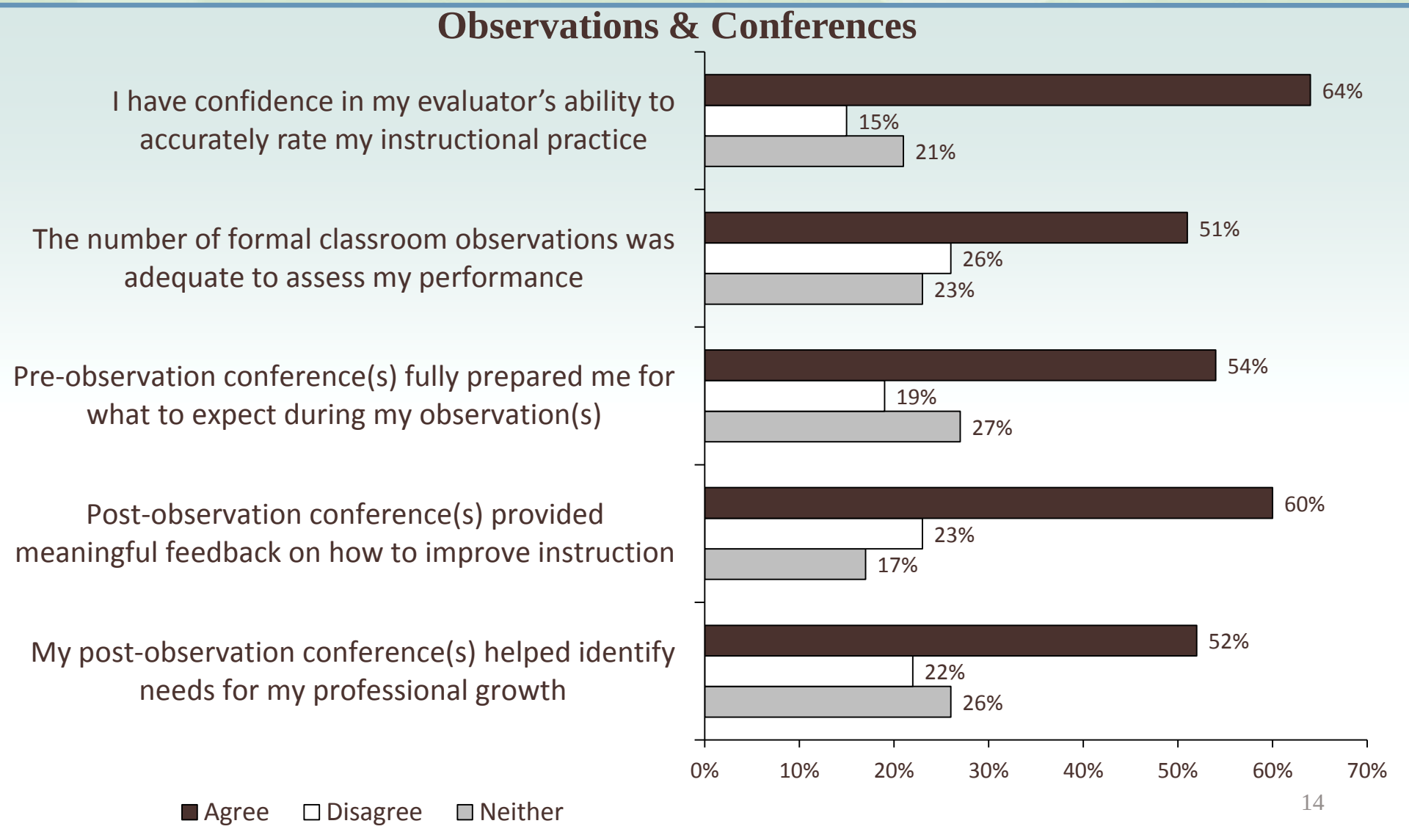
Preliminary Findings: Teaching Performance Component

- Perceptions of Danielson Framework-based observations were more varied in focus groups/interviews:
  - Positive aspects: Accurate, consistent, reasonable, helpful, specific, evidence-based, objective (less biased)
  - Negative aspects: Time consuming, inflexible, lack of relevant content expertise among observers, too easy to prepare for/script/manipulate (need for additional informal observations)

Preliminary Findings: Logistics of Observations/Conferences

- 83% of responding teachers reported 2-3 formal classroom observations
  - But informal observations varied:
    - One informal observation: 30%
    - 2-3 informal observations: 36%
    - 4+ informal observations: 34%
- 71% of respondents reported 2-3 pre-conferences, 63% reported having 2-3 post-conferences
  - These conferences generally spanned less than an hour:
    - < 15 mins (20%)
    - 15-30 mins (45%)
    - 31-45 mins (25%)

Preliminary Results: End-of-Year Surveys of Teachers



Preliminary Findings: Survey Component

- Peer review, student surveys & parent surveys were a major concern in focus groups/interviews at all pilot LEAs.
  - Common student/parent survey concerns:
    - Logistical difficulties with administration (computer/Internet access)
    - Reliability/validity of results: low response rates, overly subjective, problems "assigning" students/teachers, age/maturity of students
  - Common peer review concerns:
    - Confidentiality (some printed forms)
    - Some questions were difficult to answer knowledgeably (e.g., professional organizations), little useful feedback
    - Some reviewers assigned by principals, others picked by teachers
    - Collegiality issues/tensions/discomfort with process

Teacher Survey Results:

Survey Component

- Surveyed teachers were pessimistic about student & parent surveys providing an accurate assessment of their teaching performance.
  - Proportions of teachers responding that the following can assess their performance with moderate/high accuracy:
    - Student surveys: 50%
    - Parent surveys: 46%
- More optimism expressed about the potential for peer review:
  - 69% of surveyed teachers indicated that peer teacher surveys
    can provide an accurate assessment of their performance.

Teacher Survey Results:

Student Academic Progress Component

- Surveyed teachers were generally optimistic that student test data can provide an accurate assessment of their teaching performance.
  - Proportions of teachers responding that the following can assess their performance with moderate/high accuracy:
    - Student Learning Objectives (SLOs) established through consultation with principal: 64%
    - Standardized school-wide test scores: 60%
    - Standardized test scores from their classroom(s) of students this year: 57%

Focus Group Results:

Student Academic Progress Component

- Focus group teachers expressed confusion/concerns about student achievement component, citing:
  - Fairness issues between Group A/B teachers (different criteria) and/or those with differing student populations
  - Use of lagged test data
  - Potential for new teachers to receive higher ratings (no prior data)
- Some principals unsure about how to interpret/use rating tables (though discussions with ADE staff helped)

Other Preliminary Findings from Focus Groups

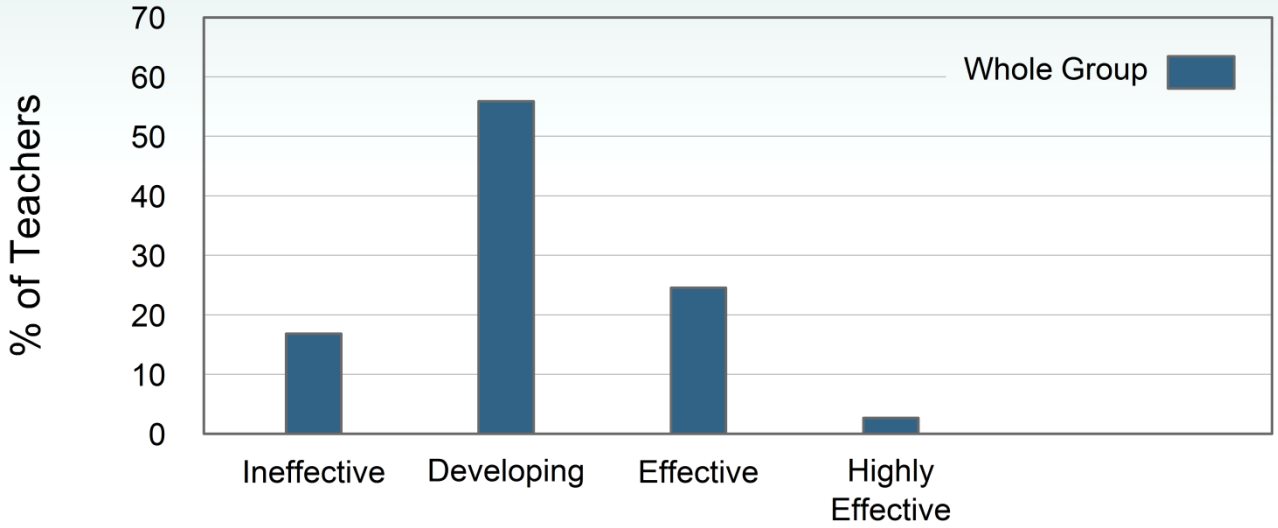
- To many principals, Danielson Framework provided clear definition of effective teaching & what to look for:
  - Principals thought conversations with teachers were more focused & in-depth (thanks to the Framework) & collaborative, particularly after pre-conferences.
  - But most focus group teachers reported that the new process had not changed conversations with their administrators
- Some participants felt overwhelmed, noting stress & agitation among teachers at the site, particularly after summative ratings were shared.

Preliminary Findings: Summative Conferences & Ratings

- 55% of responding teachers agreed that summative performance classification accurately reflected their overall performance
- In focus groups, the fairness & accuracy of final ratings was a common topic
  - Some principals expressed concerns that certain components unfairly pulled teachers' ratings down
  - Teachers expressed concerns about the difficulty of achieving a "Distinguished" rating on the Danielson Framework

Teacher Ratings by Performance Classification

Teacher Performance Classifications	
Highly Effective	108-120 points
Effective	84-107 points
Developing	60-83 points
Ineffective	<60 points



- Large number of teachers classified as "developing" or "ineffective"
- Distribution has appearance of being "skewed" to the left

Year 1 Quantitative Analysis Overview

I. Analysis of Composite Performance Scores

- Analyzed and generated any recommended revisions to scoring formulas and performance cutoff points
- Established whether the system treats any of the groups preferentially

II. Analysis of Component Observation, Survey and Student Academic Progress Data

- Established effectiveness of components in differentiating between high and low performing teachers
- Established correlational relationships between each component

III. Comparative Analysis of the Observational Instrument

- Established effectiveness and consistency of observational instrument as implemented in pilot compared to FFT data collected during the MET project

Preliminary Results Summary (Quantitative Analysis)

I. Analysis of Composite Performance Scores

– Most teachers are rated less than "Effective"
– A 0 value for student surveys skewed scores lower
– Some inconsistency in how different groups are evaluated

II. Analysis of Component Observation, Survey and Student Academic Progress Data

- Survey scores are inconsistent – large number of zero values for Student Survey
- Little consistency between classroom observations and other components

III. Item-Level Analysis of the Observational Instrument

– Appears to be biased upward and with lower variability compared to MET

– Note: Quantitative Analyses from Year 1 data will continue through SY 13-14

Changes for Project Year 2

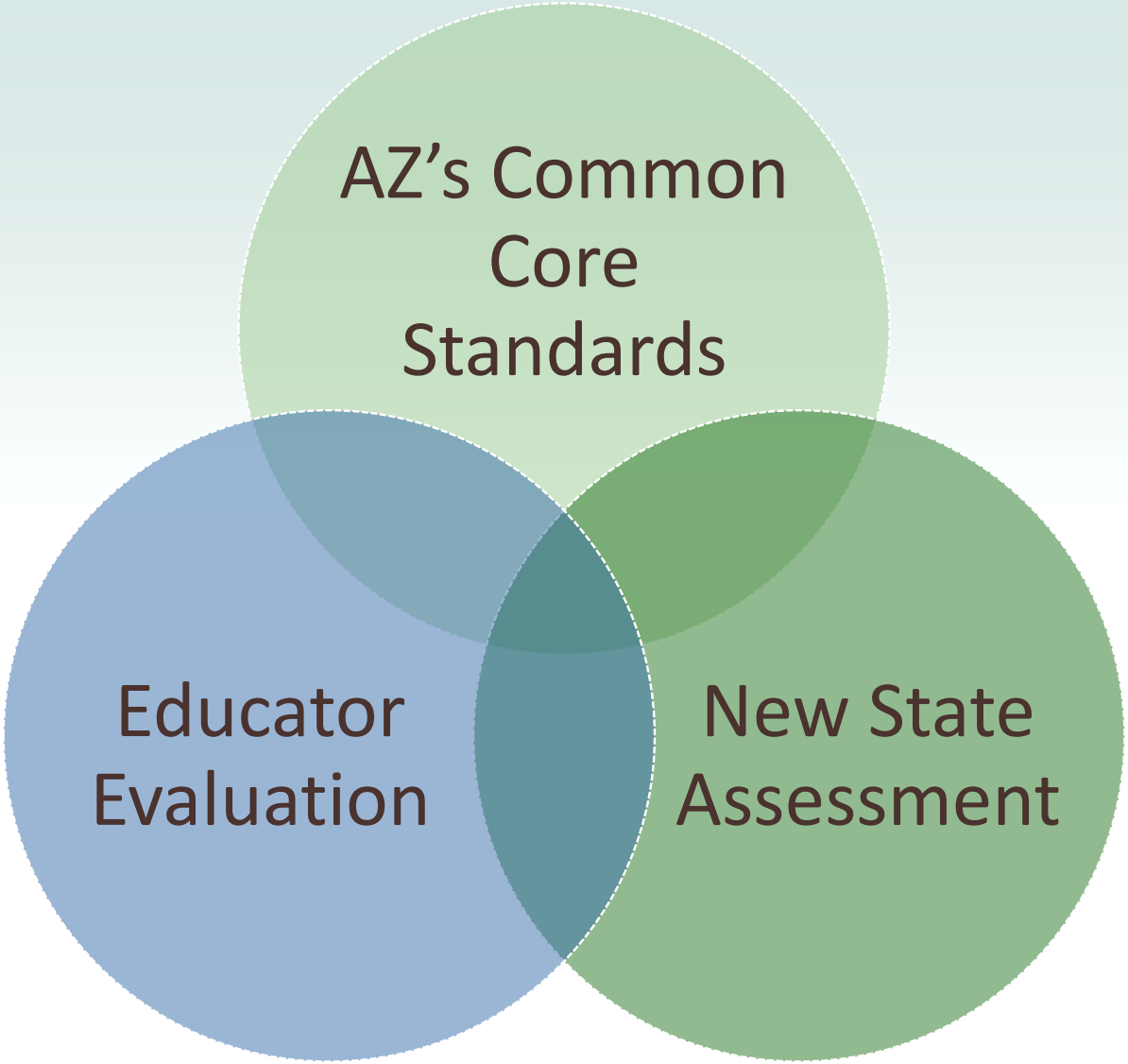
The Arizona Department of Education (ADE) has made process and implementation changes for Year 2 of the project based upon the findings extracted from the preliminary report, including:

- Streamlined Rating Tables that measure student academic progress
- Enhanced procedural directions and guidance for survey deployment
- Updated instruments to clarify language and to reflect latest legislation/requirements
- Developed “Principal Resource Guide” and other supporting guidance documents to assist with implementation
- Created comprehensive handbook on Student Learning Objectives process
- Continue to analyze cut scores and point allocations based on quantitative analysis

Project Year 2 Research Topics

1. Correlation with summative teacher performance classifications with the school's/LEA's A-F labels and AIMS data;
2. Implementation impact of including SLOs for all teachers in relation to equity, relevance and impact tied to Student Academic Progress data, growth;
3. Extent that evaluation outcomes are driving professional learning opportunities;
4. Is the ADE model changing teacher practice? Did school culture/climate perception change from Year 1?
5. Support and resources necessary for principals to be effective instructional leaders;
6. Have implementation challenges been minimized after Year 1 modifications - did processes go smoother in Year 2?

Connecting the Reform Dots



Conclusion

Any Questions?

Thank you